Supporting Information

Supramolecular protein cages constructed from a crystalline protein matrix

Hashiru Negishi[a], Satoshi Abe[a]*, Keitaro Yamashita[b], Kunio Hirata[b,c], Kento Niwase[a], Marion Boudes[d], Fasséli Coulibaly[d], Hajime Mori[e], and Takafumi Ueno[a]*

[a]School of Life Science and Technology, Tokyo Institute of Technology, Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan

[b]SR Life Science Instrumentation Unit, RIKEN/SPring-8 Center, 1-1-1, Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan

[c]Japan Science and Technology Agency, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

[d]Infection and Immunity Program, Monash Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology, Monash University, Melbourne, VIC3800, Australia

[e]Insect Biomedical Research Center, Kyoto Institute of Technology, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan

General Information

Materials

The *Spodoptera frugiperda* cell line IPLB–Sf21–AE (Sf21) was maintained in tissue culture flasks in Grace's medium (Gibco–BRL) at 27˚C with 10% fetal bovine serum (MP Biomedicals. Inc.), 2.6 mg/mL tryptose broth, 100 U/mL penicillin, and 100 μg/mL streptomycin. BaculoGold^{TM}-linealized baculovirus DNA was purchased from BD Biosciences Pharmingen. Gateway® LR Clonase® Enzyme Mix, Lipofectin® Transfer Reagent, and Top 10 Competent Cells were purchased from Life Technologies. Another reagents were purchased from TCI, Wako, Nacalai Tesque, Sigma–Aldrich and Life Technologies and were used without further purification.

Preparation of cysteine mutants of PhC

The plasmid of **1-PhM**, **2-PhM**, **3-PhM** were prepared by QuikChange Site-Directed Mutagenesis Kit (Strategene) with primers below. The plasmids were transformed into Top 10 Competent Cells and purified by QIAprep Spin Miniprep kit (Qiagen).

E73C	F: 5'-GGCACTTAAGGAATATCGCTGCGGACAACACAACGACTC-3'
	R: 5'-GAGTCGTTGTGTTGTCCGCAGCGATATTCCTTAAGTGCC-3'
Y83C	F: 5'-CGACTCTTACGATGAGTGCGAAGTGAATCAGAGCATC-3'
	R: 5'-GATGCTCTGATTCACTTCGCACTCATCGTAAGAGTCG-3'
S193C	F: 5'-CATCATACAATGTCGGATGCGCACATAACGTAATGGACG-3'
	R: 5'-CGTCCATTACGTTATGTGCGCATCCGACATTGTATGATG-3'
A194C	F: 5'-CATCATACAATGTCGGATGCTGCCATAACGTAATGGACGTCTTC-3'
	R: 5'-GAAGACGTCCATTACGTTATGGCAGCATCCGACATTGTATGATG-3'

The purified mutant plasmids were then transferred to 'destination vector' by LR reactions with the LR Clonase Enzyme Mix and purified by QIAprep Spin Miniprep kit (Qiagen). The mutant plasmids

containing baculovirus promoter were co-transfected into Sf21 cells with linearized baculovirus DNA using lipofectin reagent and incubated at 27 °C for 24 h with Grace's medium. After 24 h, Grace's medium was changed to Grace's medium with 10% fetal bovine serum and incubated at 27°C for 4 days. The supernatant of medium contain the mutant baculovirus. Expression and purification of mutants were performed by the same procedures with WTPhC.[1,2]

Crystal Structure Analysis.

Prior to the data collection, **1-PhC**, **2-PhC**, **3-PhC**, **ox1-PhC**, **ox2-PhC**, and **ox3-PhC** were immersed in a buffer solution containing 50% (w/w) ethylene glycol and were spread on MicroMesh and subsequently frozen in liquid nitrogen. X-ray diffraction data of **1-PhC**, **2-PhC**, **3-PhC**, **ox1-PhC**, **ox2-PhC** and **ox3-PhC** were collected at 100 K at beamlines BL32XU and BL41XU at SPring-8 using X-ray wavelength of 1.00 Å. The complete sets of structure factor amplitudes were obtained by merging multiple small-wedge (5° or 10° each) datasets collected from single crystals. The crystal positions in a cryoloop were identified by low-dose raster scan. At BL32XU, the whole data collection process was automated by ZOO system including sample exchange by a robot. Collected datasets were automatically processed and merged by KAMO.[3] Each dataset was indexed and integrated using XDS.[4] The datasets consistently indexed with the known cell parameter (a ~ 103 Å, $I23$) were selected and two possible reindex operators (*hkl* and -*hlk*) were tested to give better match to the previously solved data (2OH6). The datasets were subjected to hierarchical clustering by pairwise correlation coefficient of intensities. The datasets in each cluster were scaled and merged using XSCALE[4] with outlier rejections implemented in KAMO. The clusters with the highest $CC_{1/2}$ were chosen for downstream analyses. The structure was solved by rigid body refinement with phenix.refine[5] using the previously solved structure (2OH6). Refinement of the protein structure was performed at resolutions of 1.72, 1.70, 1.79, 1.65, 1.67 and 1.85 Å for **1-PhC**, **2-PhC**, **3-PhC**, **ox1-PhC**, **ox2-PhC** and **ox3-PhC**, respectively, using REFMAC5[6] in the CCP4 suite. Rebuilding was performed using COOT[7] based on sigma-A weighted

(2*Fo-Fc)* and (*Fo-Fc*) electron density maps. N-terminal region (ACE1-Thr7) and loop region (His76-Asp78) in **1-PhC**, N-terminal region (ACE1-Ser8) and loop region (Gln75-Asp78) in **2-PhC**, N-terminal region (ACE1-Thr7) in **3-PhC**, N-terminal region (ACE1-Asn9) and loop region (His76-Asp78) for **ox1-PhC**, N-terminal region (ACE1-Val4) in **ox2-PhC**, N-terminal region (ACE1-Thr7) in **ox3-PhC** could not be modeled because electron densities corresponding to these residues are missing. In addition, Glu75 was replaced to Ala in **1-PhC** and **ox1-PhC** because of low electron density of Glu75. The models were subjected to quality analysis during the various refinement stages with omit maps and RAMPAGE.[8] The diffraction and refinement statistics are summarized in Table S1 and S2, respectively.

Accession Codes: Atomic coordinates for **1-PhC**, **2-PhC**, **3-PhC**, **ox1-PhC**, **ox2-PhC** and **ox3-PhC** have been deposited in the Protein Data Bank under accession codes 5YR1, 5YR9, 5YRA, 5YRB, 5YRC and 5YRD, respectively.

Construction of protein cages from PhCs

PhCs (6 x 10^7 crystals) were soaked in 200 µL milliQ containing 100 mM H_2O_2 at 37 °C for 12 h for formation of disulfide bonds. The oxidized crystals were then washed with milliQ (3×200 µL). Then, the crystals were dissolved by soaking of the crystals in 100 mM Glycine-NaOH buffer (pH 10, 1 mL) at 25 °C for 6 h. The resulting solution was centrifuged and the supernatant were dialyzed against 20 mM Tris/HCl buffer at 4 °C for 12 h. After dialysis, the resulting solution was centrifuged and the supernatant was characterized.

SEC: The supernatant solution was eluted with size exclusion column (Superdex200, AKTA) using 20 mM Tris/HCl (pH 9) as elution buffer. Elution of proteins was monitored by absorption at 280 nm.

MALDI-TOF analysis: The sample solution were mixed with saturated solution of sinapic acid in TA solution (water:acetonitrile = 1:1 containing 0.1 wt% trifluoroacetic acid). The analysis was performed on UltrafleXtreme (Bruker).

Gel electrophoresis: The sample solution was mixed with Bromophenol blue solution by 1:1. The sample was characterized by 7.5% acrylamide gel for 90 min, 20 mA. The proteins were detected by

CBB Stain One (Nacalai Tesque) for 1 h. Stained polyacrylamide gel was treated by milliQ water for destaining.

Transmission electron microscopy (TEM): TEM analysis was performed on JEOL JEM-1400Plus after treatment of 1μM samples with 50 % methylamine tungstate for negative stain.

Atomic force microscopy (AFM): AFM was performed on MFP 3D by Oxford Instruments Asylum Research. BL-AC40TS-C2 (Olympus) was used as cantilever tip. The samples (70 nM) were loaded into freshly cleaved mica surface for 5 h and washed with 20 mM Tris/HCl (pH 9). The images were measured in 20 mM Tris/HCl (pH 9.0).

Circular dichroism (CD) spectroscopy: CD spectra of **ox1-PhC** based cage and R13APhM were measured at 20 °C in 20 mM Tris/HCl (pH9.0) ([**ox1-PhC**-based cage] = 1.7 μM, [R13APhM] = 5 μM by J-820 (JASCO). The CD spectra data were analyzed with BeStSel server for secondary structure determination.[9] The secondary structure contents of helix and β-sheet in the crystal structure of **ox1-PhC** were estimated by DSSP program.[10]

Figure S1. Microscopy images of the **1-PhC**, **2-PhC** and **3-PhC** in Sf21 cells.

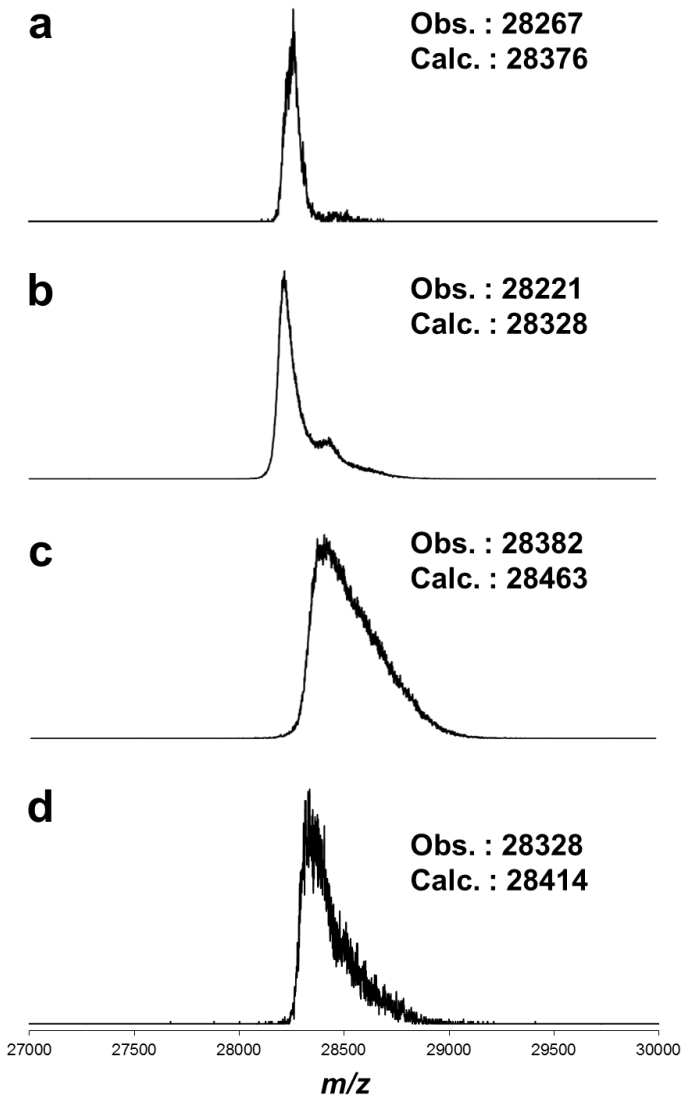

Figure S2. MALDI TOF mass spectra of (a) **1-PhC**, (b) **2-PhC**, (c) **3-PhC** and (d) R13APhC.

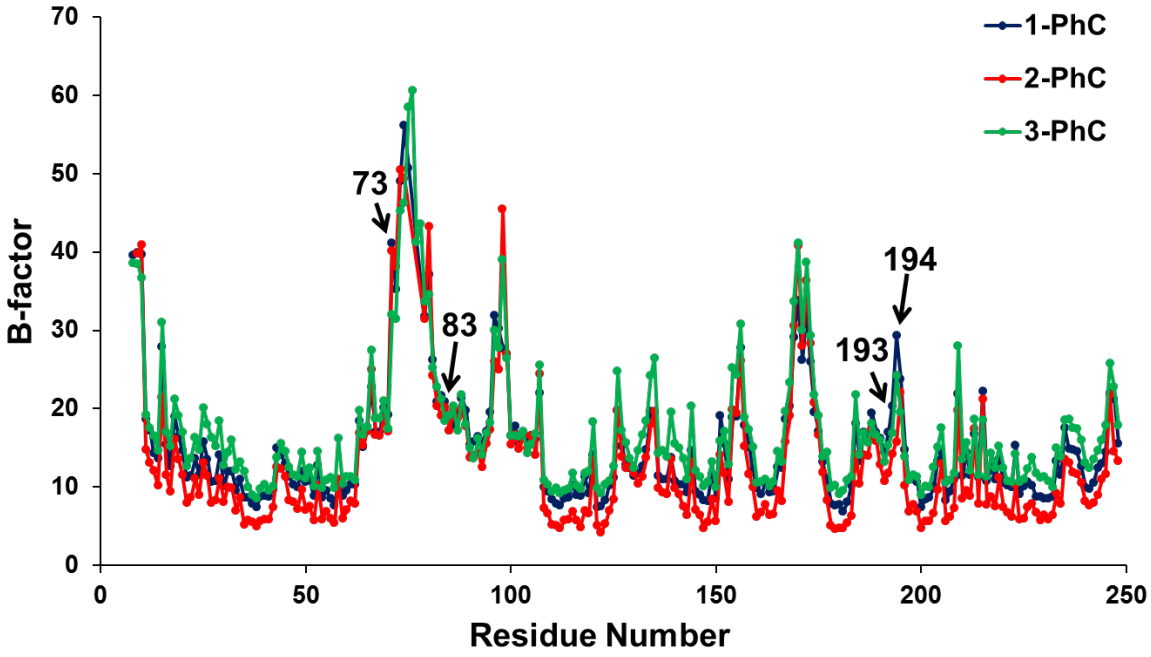

Figure S3. Average *B*-factors for main chain atoms of **1-PhC**, **2-PhC** and **3-PhC**.

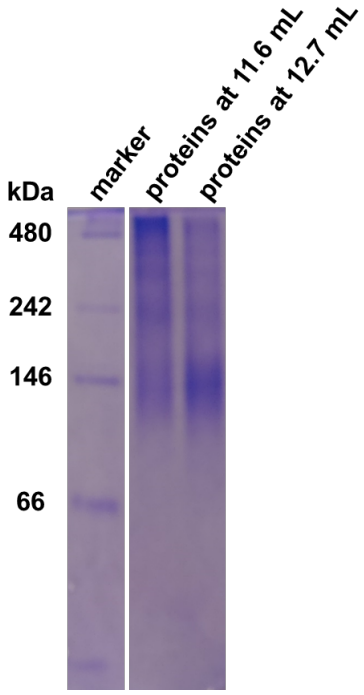

Figure S4. A polyacrylamide gel electrophoresis (PAGE) of the protein assemblies eluted at 11.6 mL and 12.7 mL from **ox1-PhC** under native conditions.

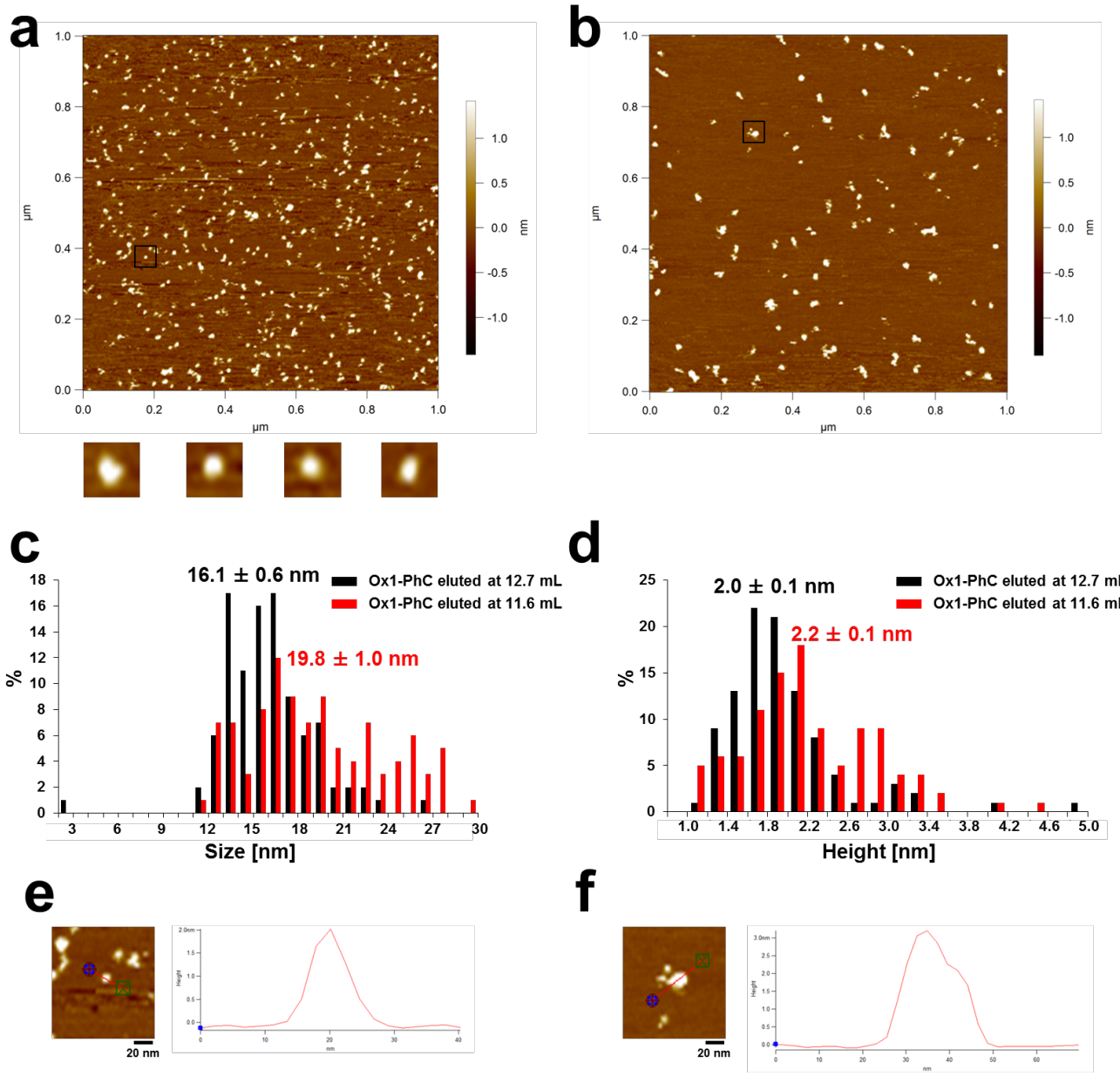

Figure S5. AFM images of (a) protein cages eluted at 12.7 mL and (b) protein aggregations eluted at 11.6 mL from **ox1-PhC**. (c) Size distribution of protein cages from **ox1-PhC** (N_{cage} = 100, N_{agg} = 105). (d) Height distribution of protein cages from **ox1-PhC** (N_{cage} = 100, N_{agg} = 105). (e, f) Height profiles of (e) **ox1-PhC**-based cage and (f) aggregation were measured along red lines of the assemblies in (e) **ox1-PhC**-based cage and (f) aggregation.

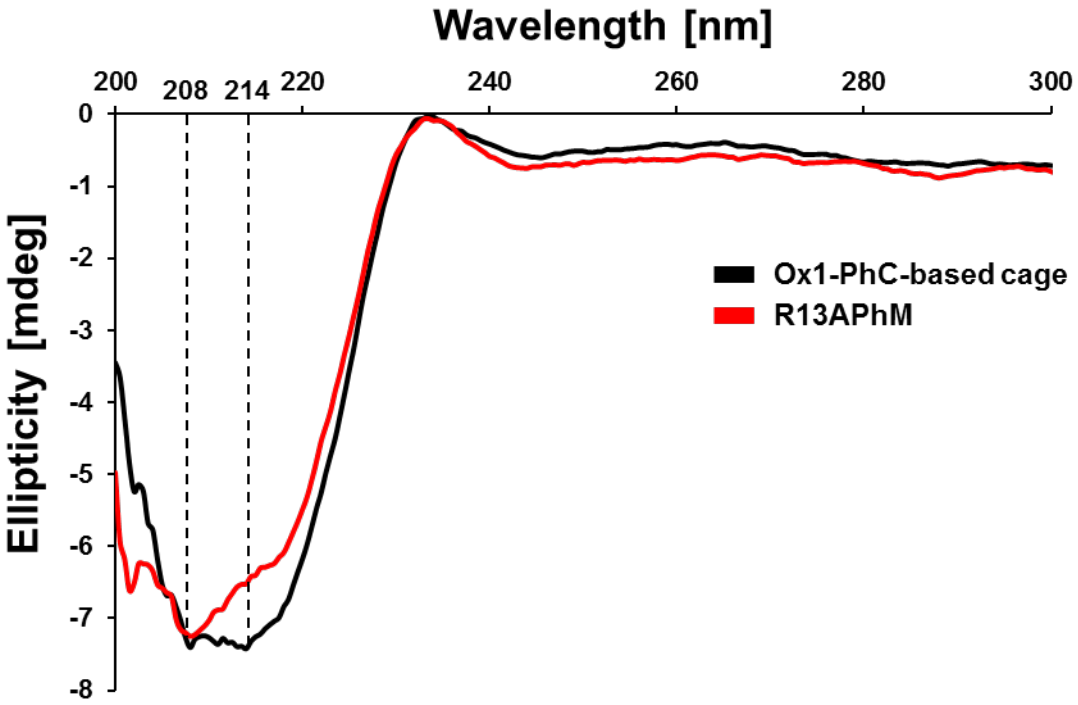

Figure S6. CD spectroscopy of **ox1-PhC**-based cages (black line) and R13APhM (red line). Each CD spectrum of sample was measured at 20 °C ([**ox1-PhC**-based cage] = 1.7 μM, [R13APhM] = 5 μM).

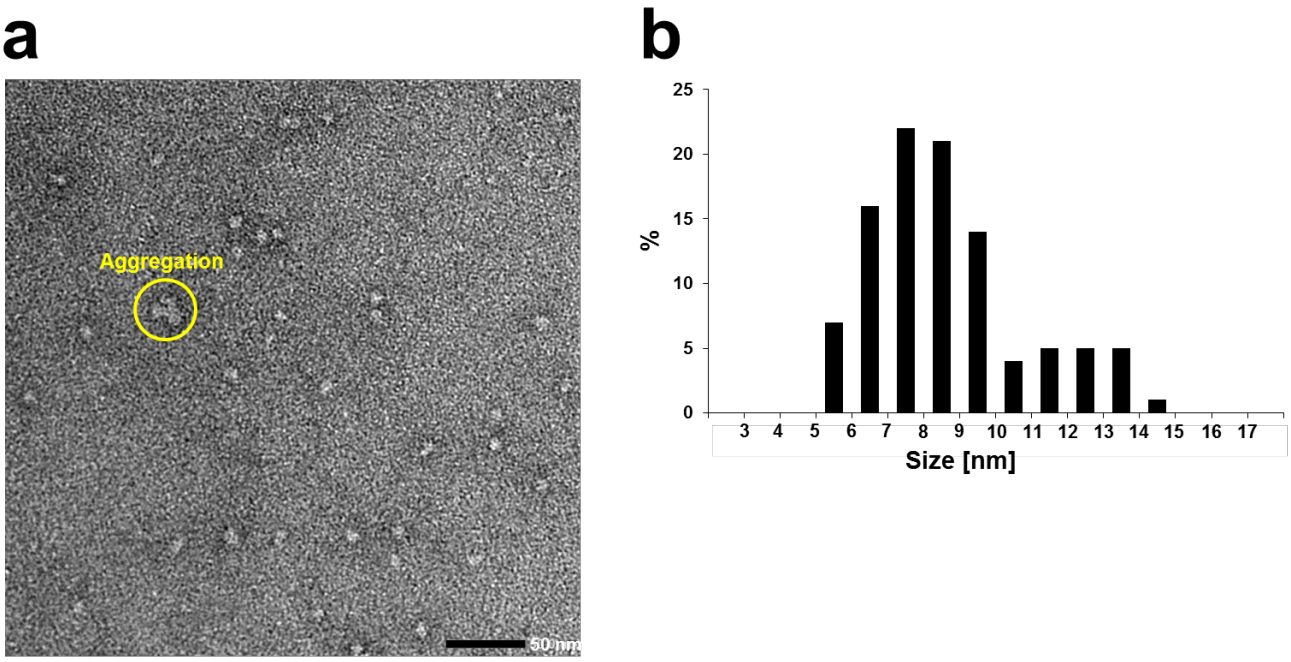

Figure S7. (a) A TEM image of protein aggregations eluted by SEC at 11.6 mL from **ox1-PhC**. (b) Size distribution of protein aggregations from **ox1-PhC** in TEM images (n = 101).

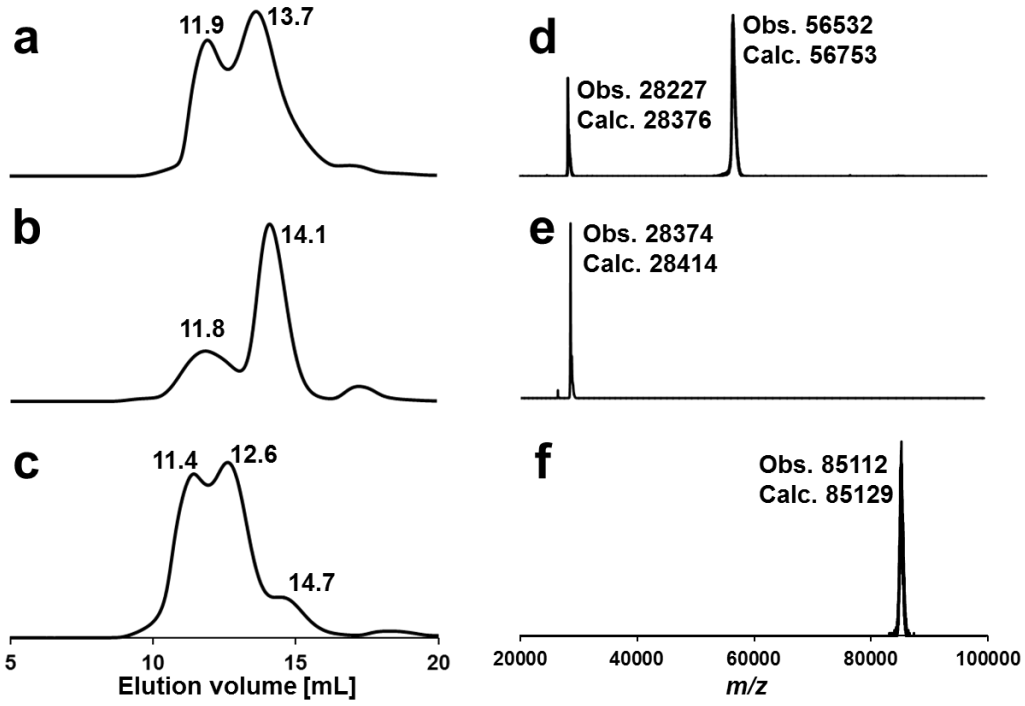

Figure S8. (a-c) SEC elution profiles of protein assemblies extricated from (a) **1-PhC**, (b) **oxR13APhC** and (c) **ox1-PhC** after air-drying. (d-f) MALDI-TOF MS of eluted protein assemblies. (d) Protein assemblies eluted at 13.7 mL from **1-PhC**. (e) Protein assemblies eluted at 14.1 mL from **oxR13APhC**. (f) Protein assemblies eluted at 12.6 mL from **ox1-PhC** after air-drying.

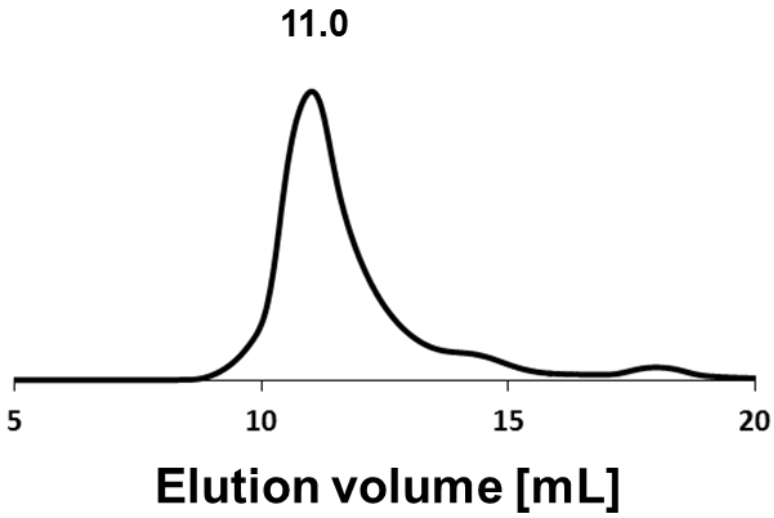

Figure S9. SEC profile of oxidized **1-PhM** in solution

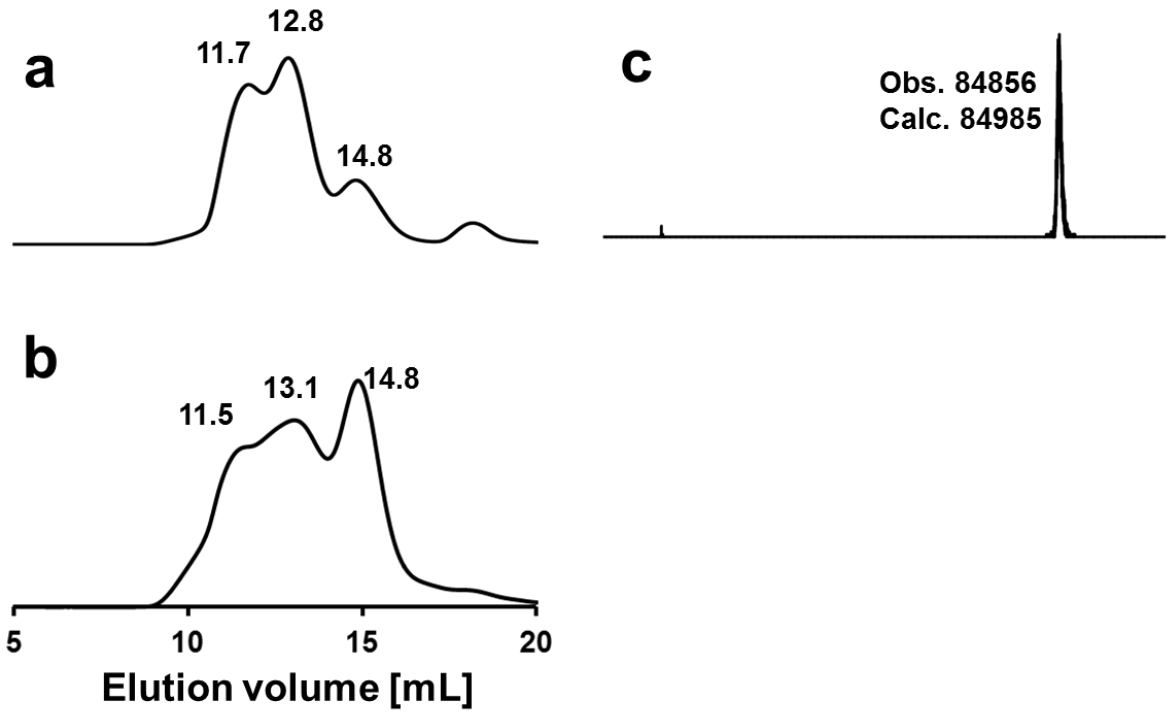

Figure S10. Characterization of protein assemblies from PhCs by SEC and MALDI TOF MS. (a, b) SEC elution profiles of protein assemblies from (a) **ox2-PhC** and (b) **ox3-PhC**. (c) MALDI TOF mass spectrum of protein assemblies eluted at 12.8 mL from **ox2-PhCs**.

Table S1. Crystallographic data

Data collection	**1-PhC**	**2-PhC**	**3-PhC**	**ox1-PhC**	**ox2-PhC**	**ox3-PhC**
No. collected datasets	32	127	49	78	323	29
No. of merged datasets	9	51	30	13	119	17
Space group	*I*23	*I*23	*I*23	*I*23	*I*23	*I*23
Crystal cell (Å)						
$a = b = c$	103.6	103.7	103.7	103.7	103.7	103.6
Resolution range (Å)	50.0–1.72 (1.82–1.72)	50.0–1.70 (1.80–1.70)	50.0–1.79 (1.90–1.79)	50.0–1.65 (1.75–1.65)	50.0–1.67 (1.77–1.67)	50.0–1.85 (1.96–1.85)
Completeness (%)	98.5 (99.2)	100.0 (100)	100.0 (100)	99.7 (99.8)	100.0 (100)	99.9 (99.8)
Multiplicity	4.9 (4.9)	27.3 (18.8)	21.1 (20.4)	7.0 (7.1)	60.5 (36.5)	9.0 (8.5)
Unique reflections	19520 (3161)	20579 (3282)	17636 (2836)	22401 (3581)	21659 (3432)	15953 (2542)
R_{meas}	0.244 (1.132)	0.599 (6.046)	0.579 (3.072)	0.204 (1.431)	0.631 (6.329)	0.402 (1.845)
I/σ	5.57 (1.34)	8.49 (1.10)	6.51 (1.00)	7.37 (1.34)	11.24 (1.21)	5.52 (1.21)
CC1/2	0.984 (0.521)	0.990 (0.507)	0.988 (0.539)	0.992 (0.498)	0.994 (0.563)	0.980 (0.499)

Values in parentheses are for the highest-resolution shell. Friedel pairs are treated as different reflections.

Table S2. Refinement statistics.

	1-PhC	**2-PhC**	**3-PhC**	**ox1-PhC**	**ox2-PhC**	**ox3-PhC**
Resolution range (Å)	32.76-1.72	36.68-1.70	42.34-1.79	36.67-1.65	36.66-1.67	42.30-1.85
Reflection used	17570	18547	15866	20176	19524	14365
R-factor (%)	15.61	15.79	15.84	15.65	14.27	16.20
Free *R*-factor (%)	20.85	20.47	21.78	20.15	18.57	22.18
R.m.s. deviations from ideal						
Bond length (Å)	0.0173	0.0195	0.0176	0.0193	0.0195	0.0156
Angle (°)	1.7803	1.8630	1.7536	1.9107	1.9011	1.6847
Ramachandran plot (%)						
most favored	96.09	96.43	96.62	96.93	97.05	96.64
allowed	3.91	3.57	2.95	2.63	2.95	2.52

References

1. H. Mori, R. Ito, H. Nakazawa, M. Sumida, F. Matsubara and Y. Minobe, *J. Gen. Virol.*, 1993, **74**, 99-102.
2. S. Abe, H. Tabe, H. Ijiri, K. Yamashita, K. Hirata, K. Atsumi, T. Shimoi, M. Akai, H. Mori, S. Kitagawa and T. Ueno, *Acs Nano*, 2017, **11**, 2410-2419.
3. https://github.com/keitaroyam/yamtbx/blob/master/doc/kamo-en.md.
4. W. Kabsch, *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, 2010, **66**, 125-132.
5. P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart and P. D. Adams, *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, 2012, **68**, 352-367.
6. G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Crystallogr., Sect. D*, 1997, **53**, 240-255.
7. P. Emsley and K. Cowtan, *Acta Crystallogr., Sect. D*, 2004, **60**, 2126-2132.
8. S. C. Lovell, I. W. Davis, W. B. Adrendall, P. I. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson and D. C. Richardson, *Proteins Struct. Funct. Genet.*, 2003, **50**, 437-450.
9. A. Micsonai, F. Wien, L. Kernya, Y. H. Lee, Y, Goto, M. Refregiers and J. Kardos, *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112*, E3095-E3103.
10. W. Kabsch and C. Sander, *Biopolymers* **1983,** *22*, 2577-2637.